

ABSTRACT

The present invention provides mathematical model-based incremental clustering methods for classifying sets of data and predicting new data values, based upon the concepts of similarity and cohesion. In order to increase processing efficiency, these methods employ weighted attribute relevance in building unbiased classification trees and sum pairing to reduce the number of nodes visited when performing classification or prediction operations. In order to increase prediction accuracy, these methods employ weighted voting over each value of target attributes to calculate a prediction profile. The present invention also allows an operator to determine the importance of attributes and reconstitute classification trees without those attributes deemed unimportant to further increase classification structure node processing efficiency. An operator can vary instance attribute values via a graphical user interface to explore the domain space of a classified data set, and use the visualization aspect of the present invention to visually contrast data set members with distinguishing features. Federated classification and prediction operations are also provided by the present invention to provide enhanced scalability in a multiprocessing and/or parallel processing environment. Additionally, the sum pairing of values, federation and multiprocessing/parallel processing enable the present invention to perform classification and prediction in real-time.